Engineering a Thermostable Protein via Optimization of Charge−Charge Interactions on the Protein Surface[†]

Vakhtang V. Loladze,[‡] Beatriz Ibarra-Molero,[§,||] Jose M. Sanchez-Ruiz,*,[§] and George I. Makhatadze*,[‡]

*Department of Biochemistry and Molecular Biology, Pennsylvania State College of Medicine,
Hershey, Pennsylvania 17033-0850, and Departamento de Quimica Fisica, Facultad de Ciencias,
Universidad de Granada, 18071 Granada, Spain*

ABSTRACT: A simple theoretical model for increasing the protein stability by adequately redesigning the distribution of charged residues on the surface of the native protein was tested experimentally. Using the molecule of ubiquitin as a model system, we predicted possible amino acid substitutions on the surface of this protein which would lead to an increase in its stability. Experimental validation for this prediction was achieved by measuring the stabilities of single-site-substituted ubiquitin variants using urea-induced unfolding monitored by far-UV CD spectroscopy. We show that the generated variants of ubiquitin are indeed more stable than the wild-type protein, in qualitative agreement with the theoretical prediction. As a positive control, theoretical predictions for destabilizing amino acid substitutions on the surface of the ubiquitin molecule were considered as well. These predictions were also tested experimentally using correspondingly designed variants of ubiquitin. We found that these variants are less stable than the wild-type protein, again in agreement with the theoretical prediction. These observations provide guidelines for rational design of more stable proteins and suggest a possible mechanism of structural stability of proteins from thermophilic organisms.

How to stabilize protein structure? This question is not just of scholastic interest. The answer to this question has immediate importance for the biotechnological industry, which is interested in improving the thermostability of enzymes. The efforts in this direction have concentrated on repacking the hydrophobic cores, engineering disulfide bridges, adding extra hydrogen bonds or salt bridges, and improving secondary structure propensities or side chain helix dipole interactions (*1−8*). All these methods can be characterized as optimizing the local short-range interactions. Such an approach is well justified because it is becoming more and more clear that the protein folding is a hierarchical process and thus is mostly driven by local interactions (*9, 10*). However, this does not mean that the final native state of the protein is stabilized exclusively by the short-range interactions (*11, 12*). Long-range interactions such as charge− charge interactions will also contribute to the stability of proteins. To address this issue, we developed a simple approach [the "TK−BK procedure" (*13*)] that allows us, nevertheless, to determine the regions of the protein surface where redesign

of the charge−charge interactions is likely to lead to stability enhancement. In the TK−BK procedure (*13*), the interaction energies between unit positive charges placed in the protonation sites of ionizable groups are estimated using the solvent-accessibility-corrected Tanford−Kirkwood model (*14*), and subsequently, the pairwise charge−charge interaction energies are calculated as averages over the protonation states of the native protein on the basis of the Bashford− Karplus reduced-set-of-sites approximation (*15*). This approach is indeed a simple one (charge−charge interactions in the unfolded state and solvation terms in the native state are neglected), but precisely because of its simplicity, it leads to easily testable predictions. Thus, one of the outcomes of our TK−BK calculation is the energy value $\langle W_i \rangle$ for each ionizable group from the charge−charge interaction of group *i* with the rest of the ionizable groups in the protein. A positive value for $\langle W_i \rangle$ indicates that the group is predominantly involved in destabilizing interactions with groups with the same charge and, accordingly, that charge deletion mutations at position *i* will eliminate destabilizing interactions and enhance the stability of the protein. Even larger stability enhancements could in principle be obtained by charge reversal mutations, since charge reversal will change the sign of the interactions involving position *i*, thus turning the destabilizing interactions into stabilizing ones. To summarize, we suggest that groups with positive $\langle W_i \rangle$ values should be selected for charge deletion or charge reversal mutation, provided, of course, that some obvious requirements are met; for instance, mutations should be carried out at well-exposed positions, since the Tanford−Kirkwood model does not take into account solvation contributions.

* To whom correspondence should be addressed. Fax: (717) 531-7072 (G.I.M.) or 34-958-272879 (J.M.S.-R.). E-mail: makhatadze@psu.edu (G.I.M.) or sanchezr@goliat.ugr.es (J.M.S.-R.).
[‡] Pennsylvania State College of Medicine.
[§] Universidad de Granada.
[||] Present address: Department of Chemistry, The Pennsylvania State University, University Park, PA 16802.

To validate this approach experimentally, we applied the TK−BK procedure to the molecule of ubiquitin. Ubiquitin is a small globular protein with 76 amino acid residues. The secondary structure of this protein consists of a three-full turn α-helix positioned on top of a four-strand β-sheet (*16*). Unfolding of the ubiquitin was shown to be a two-state process in both the equilibrium (*17*, *18*) and kinetic experiments (*19*). The molecule has an almost equal number of basic (four arginines, seven lysines, and one histidine) and acidic residues (six aspartates and five glutamates) which all are highly solvent exposed. This suggests (*13*, *20*) that the effect of the local environment on the ionization properties of residues in the native state will be greatly reduced. We generated 11 single-site amino acid substitutions of the ubiquitin molecule at the positions which according to the TK−BK procedure would lead to the changes in stability and experimentally measured the stabilities of these ubiquitin variants. We show that there is excellent qualitative agreement between experimentally measured and predicted values for the stabilization of the ubiquitin molecule.

MATERIALS AND METHODS

Site-directed mutagenesis, overexpression, and purification of the ubiquitin variants were performed as described previously (*17*). Changes in ellipticity at 222 nm as a function of urea concentration were monitored on a JASCO J-715 spectropolarimeter at 25 °C using 1 mm cylindrical quartz cells. Two protein stock solutions at a concentration of 0.5 mg/mL were prepared. Solution A contained protein in 50 mM sodium acetate (pH 5.0). Solution B contained protein in 10 M urea and 50 mM sodium acetate (pH 5.0). The protein concentration in these two solutions was equalized spectrophotometrically using a Hitachi U2001 spectrophotometer (*17*, *18*). Protein solutions at varied urea concentrations were prepared by mixing solutions A and B in different proportions. Final urea concentrations were obtained by measuring the diffraction index as described previously (*21*).

Transition curves were analyzed according to a two-state model where the equilibrium constant, $K_{eq}(C)$, is given by

$$K_{eq}(C) = \frac{\Theta_N(C) - \Theta_X(C)}{\Theta_X(C) - \Theta_U(C)} \qquad (1)$$

where $\Theta_N(C)$ and $\Theta_U(C)$ are the dependencies of the ellipticity at 222 nm on denaturant concentrations (taken to be linear) for the native and unfolded states, respectively, and $\Theta_X(C)$ is the experimentally measured ellipticity at 222 nm. The Gibbs energy at a given concentration of denaturant, $\Delta G(C)$, is related to the equilibrium constant

$$\Delta G(C) = -RT \ln K_{eq}(C) \qquad (2)$$

To obtain the Gibbs energy at zero denaturant, ΔG°_{exp}, the linear extrapolation method (LEM)[1] was used (*21*, *22*):

$$\Delta G(C) = \Delta G^\circ_{exp} - mC_{1/2} \qquad (3)$$

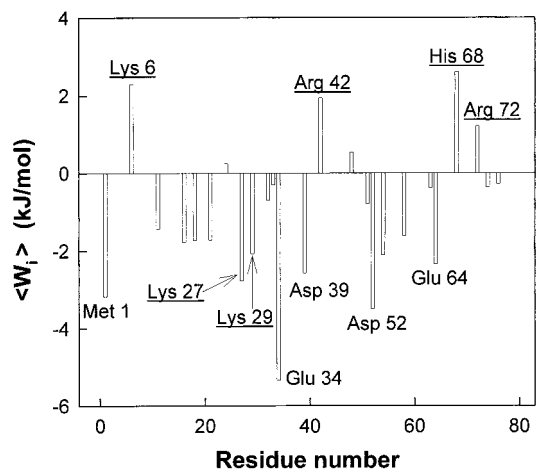where the so-called denaturant m value is the dependence

---

FIGURE 1: Bar graph of the energies due to the charge−charge interactions of all ionizable residues in the ubiquitin molecule at pH 5.0 as a function of the residue number in the sequence. Calculations have been carried out using the Tanford−Kirkwood model (*14*) with the Bashford−Karplus approximation (*15*) as described previously (*13*). Positive values of $\langle W_i \rangle$ indicate that the amino acid side chains are involved in predominantly destabilizing charge−charge interactions, while negative values of $\langle W_i \rangle$ correspond to the amino acid side chains that are involved in predominantly stabilizing interactions. Selected residues are identified above each bar. Underlined amino acid residues are the sites where substitutions were made.

of the Gibbs energy of unfolding on denaturant concentration and $C_{1/2}$ is the urea concentration at the midpoint of a two-state transition. Data for all ubiquitin variants were simultaneously analyzed (global analysis) according to eqs 1−3 using common $\Theta_N(C)$ and $\Theta_U(C)$ functions and a common m value. Errors in ΔG°_{exp} are 95% confidence intervals.

The calculations of the energy of charge−charge interactions were performed using the solvent-accessibility-corrected Tanford−Kirkwood model (*14*) with Bashford−Karplus reduced-set-of-sites approximation (*15*); this TK−BK procedure was performed as described previously (*13*).

RESULTS AND DISCUSSION

The results of TK−BK analysis of the ubiquitin molecule in terms of charge−charge interactions at pH 5.0 are presented in Figure 1. Each bar represents the energy, $\langle W_i \rangle$, of charge−charge interactions of a given residue, *i*, with all other residues, including charges at the C- and N-termini. It must be emphasized that the calculated energy term estimates only charge−charge interactions and only in the native state of the protein and it does not take into account other interactions such as hydrogen bonding, charge helix dipole interactions, secondary structure propensity, etc. It is clear from Figure 1 that most ionizable groups in ubiquitin are in favorable environments from the point of view of charge−charge interactions (i.e., most groups are predominantly interacting with charges with the opposite sign). There are, however, six residues which are predicted to have unfavorable charge−charge interactions. These are Lys6, Asp24, Arg42, Lys48, His68, and Arg72. Two of these, Asp24 and Lys48, are predicted to have relatively small unfavorable contributions, only 0.3 and 0.5 kJ/mol, respectively. For comparison, the other four residues (Lys6, Arg42, His68, and Arg72) are predicted to have destabilizing charge−charge energies that are several times larger (2.3, 2.0, 2.6,

Table 1: Structural and Thermodynamic Description of Ubiquitin Variants

ubiquitin variant	fraction exposed[a] (%)	secondary structure[b]	ΔG°_{exp}[c] (kJ/mol)	$C_{1/2}$[d] (M)	$\Delta\Delta G^\circ_{exp}$[e] (kJ/mol)	$\Delta\Delta G_{q-q}$[f] (kJ/mol)
wild type	–	–	25.2 ± 1.0	5.99	0	?
K6E	65	β-1	27.4 ± 1.1	6.51	2.2	4.5
K6Q	65	β-1	26.3 ± 1.1	6.25	1.1	2.4
K27Q	11	α	17.2 ± 0.7	4.09	−8.0	−2.6
K29Q	45	α	19.0 ± 0.8	4.51	−7.0	−2.0
K29N	45	α	18.2 ± 0.7	4.32	−6.2	−2.0
R42E	49	β-4	32.0 ± 1.3	7.60	6.8	3.6
H68Q	58	β-3	27.5 ± 1.1	6.53	2.3	2.8
H68E	58	β-3	28.4 ± 1.2	6.75	3.2	5.4
R72Q	58	β-3	23.8 ± 1.0	5.65	−1.4	1.3

[a] Fraction exposed is defined as ASA^N/ASA^U, where ASA^N and ASA^U are the water accessible surface areas of the amino acid side chain in the native and unfolded states, respectively, calculated as described previously (*28*). [b] Secondary structure of the residue as given in the header of Protein Data Bank file 1UBQ. [c] ΔG_{exp} was obtained by globally fitting the experimental data depicted in Figure 2 to eq 3 using the same *m* (4210 ± 160 J/mol²). [d] $C_{1/2}$ is the concentration of urea at which the populations of the native and unfolded molecules are the same; i.e., $\Delta G(C_{1/2}) = 0$. [e] Defined as $\Delta\Delta G^\circ_{exp} = \Delta G^\circ_{exp}(mut) - \Delta G^\circ_{exp}(WT)$. [f] Defined as $\Delta\Delta G_{q-q} = \Delta G_{q-q}(mut) - \Delta G_{q-q}(WT)$, where $\Delta G_{q-q}(WT)$ and $\Delta G_{q-q}(mut)$ were calculated using the TK−BK procedure as described previously (*13*).
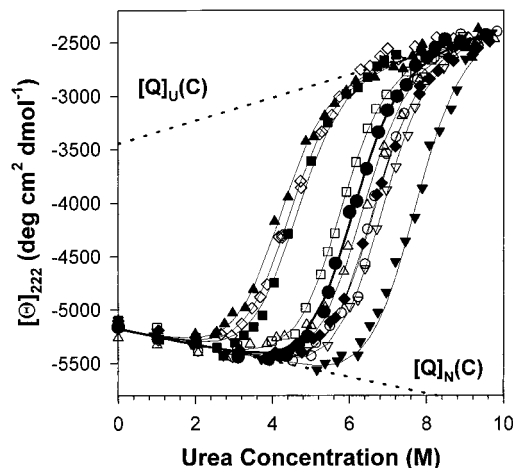

FIGURE 2: Dependence of the ellipticity at 222 nm on the urea concentration for the ubiquitin variants: wild type (●), K6E (◆), K6Q (△), K27Q (▲), K29Q (■), K29N (◇), R42E (▼), H68Q (○), H68E (▽), and R72Q (□). Dashed lines show the dependencies of ellipticities of the native, $\Theta_N(C)$, and unfolded, $\Theta_U(C)$, states on urea concentrations. Solid lines show the results of the nonlinear regression analysis according to eqs 1−3.

and 1.2 kJ/mol at pH 5.0, respectively). Then basic to neutral amino acid and basic to acidic amino acid substitutions at these four positions are expected to eliminate unfavorable charge−charge interactions or to replace them with stabilizing interactions in the case of the charge reversal substitutions, thus increasing protein stability. This prediction can be easily tested experimentally, by measuring the changes in the stability of ubiquitin variants generated using site-directed mutagenesis.

We generated six variants of ubiquitin with substitutions at residues Lys6 (K6E and K6Q), Arg42 (R42E), His68 (H68Q and H68E), and Arg72 (R72Q). The TK−BK procedure was applied to these six variants. For the variants resulting from charge reversal mutations, we assumed for calculation purposes that the new charge was in the same position and had the same accessibility as the old one. The charge−charge contributions to the unfolding Gibbs energies were obtained using the TK−BK procedure, and the results relative to the wild type (i.e., as $\Delta\Delta G_{q-q}$ values) are given in Table 1. Note that $\Delta\Delta G_{q-q} > 0$ for the six ubiquitin variants. Thus, on the basis of the calculations of charge−charge interactions in ubiquitin, all six amino acid substitutions (K6E, K6Q, R42E, H68Q, H68E, and R72Q) will lead to an *increase* in the stability of the variants, as was to be expected from the positive ⟨W_i⟩ values for the four positions where substitutions were introduced (Figure 1).

As a control, we also generated ubiquitin variants with amino acid substitutions at positions 27 and 29 (Figure 1). The energies of charge−charge interactions of Lys27 and Lys29 calculated according to the TK−BK procedure are stabilizing by 2.8 and 2.1 kJ/mol at pH 5.0, respectively. We generated three ubiquitin variants (K27Q, K29N, and K29Q), replacing basic Lys27 and Lys29 with neutral Gln or Asn residues. These variants of ubiquitin are expected to have a stability *lower* than that of the wild-type protein, as judged by the ⟨W_i⟩ values of the substituted positions (Figure 1) and the energies of charge−charge interactions relative

to that of the wild type, $\Delta\Delta G_{q-q}$, calculated on virtually substituted structures (Table 1).

The structure and stability of the ubiquitin variants were studied using far-UV CD spectroscopy. The spectra for the variants collected in the range from 250 to 195 nm had shapes and absolute values similar to those of the wild-type protein, indicating that there were no detectable structural changes upon substitution (data not shown). Such an outcome was expected on the basis of the consideration that all amino acid substitutions were introduced at the solvent-exposed positions of the ubiquitin molecule (Table 1).

The unfolding of the ubiquitin variants was induced by changing the concentration of urea in solution at pH 5.0 and was monitored by changes in the ellipticity at 222 nm. The actual results of measurements of ellipticity after correction for the protein concentration are plotted as a function of urea concentration in Figure 2. From the data presented in Figure 2, it is clear that the generated variants of ubiquitin have stabilities different from that of the wild type with the amino acid substitutions having both stabilizing and destabilizing effects on the protein stability. Quantitative analysis of the unfolding curves was performed according to a two-state model using eqs 1−3. The results are presented in Table 1. For the wild-type protein, the stability is estimated to be 25.2 ± 1.0 kJ/mol under these solvent conditions (25 °C and pH 5.0), which is in agreement with the value expected from the pH dependence of the Gibbs free energy reported by us previously (*13*). The diversities in stabilities of the ubiquitin variants are significant and range from 17.2 to 32.0 kJ/mol. This results in changes in stabilities relative to that of the wild-type protein, $\Delta\Delta G^\circ_{exp}$, ranging from −8.0 to 6.8 kJ/mol (Table 1).

The experimentally obtained changes in stabilities of the ubiquitin variants at 25 °C and pH 5.0, $\Delta\Delta G^\circ_{exp}$, are compared with the changes expected on the basis of the changes in the charge−charge interactions, $\Delta\Delta G_{q-q}$, in Figure 3. Several interesting features can be observed from the plot in this figure.

The plot of $\Delta\Delta G^\circ_{exp}$ versus $\Delta\Delta G_{q-q}$ is clearly linear with a positive slope, indicating that there is a direct correlation
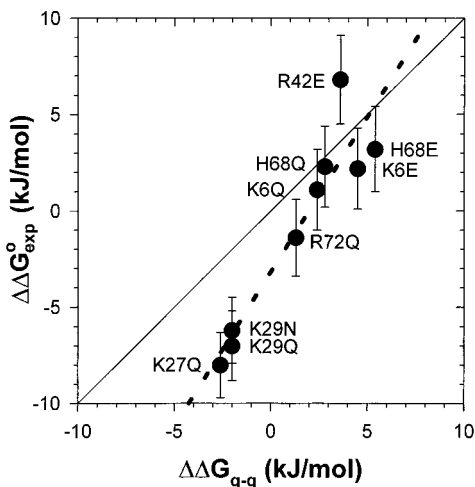
FIGURE 3: Correlation between calculated and experimentally measured changes in the Gibbs energies of ubiquitin variants with site-directed substitutions. The dashed line is the linear fit of the data with a slope of 1.6. The solid line shows the line with a slope of 1.0 expected for a perfect correlation.

between the experimental and predicted values of the stability of ubiquitin variants (correlation coefficient of 0.98). Nevertheless, the slope of the line is 1.68, significantly different from 1, and it does not pass through the origin. This means that the predicted values are only *qualitatively* tracing the experimental values of $\Delta\Delta G^{\circ}_{exp}$ for the ubiquitin variants. Such qualitative agreement between $\Delta\Delta G^{\circ}_{exp}$ and $\Delta\Delta G_{q-q}$ can be anticipated because the calculated energies account only for the charge−charge interactions and ignore all other numerous possible contributions (for a review, see refs *9* and *10* by Baldwin and Rose). More remarkable is the fact that although the predicted values are different from the experimental ones, there is a significant correlation between $\Delta\Delta G^{\circ}_{exp}$ and $\Delta\Delta G_{q-q}$. The existence of such a correlation is more important than the absolute values and can be interpreted as showing that the charge−charge interactions under these solvent conditions are the most important interactions for basic residues at the largely solvent-exposed sites of the ubiquitin molecule.

The K27Q, K29N, and K29Q variants of ubiquitin were generated to serve as a positive control. Amino acid residues Lys27 and Lys29 were predicted to have favorable stabilizing charge−charge interactions, and thus, their substitution was expected to lead to destabilization. This is exactly what was observed experimentally. K27Q, K29N, and K29Q variants of ubiquitin were found to be 8.0, 7.0, and 6.2 kJ/mol less stable than the wild-type protein, respectively. We must note that the absolute values of destabilization energies are much higher than the predicted ones (Table 1 and Figure 3), presumably due to the presence of interactions other than charge−charge interactions which were perturbed upon amino acid substitution. However, *qualitatively* amino acid substitutions at positions 27 and 29 were predicted to be destabilizing, and we observed experimentally the decrease in the stabilities of K27Q, K29N, and K29Q variants of ubiquitin.

The K6E, K6Q, H68Q, H68E, and R42E variants of ubiquitin have increased stabilities compared to that of the wild-type protein. At position 6, the stability increase was 2.2 and 1.1 kJ/mol for K6E and K6Q variants, respectively. Amino acid substitutions at position 68 led to an increase in

stability of 2.3 kJ/mol (H68Q) and 3.2 kJ/mol (H68E). The most dramatic increase in stability (6.8 kJ/mol) was observed for the R42E variant of ubiquitin. These three sites have overall unfavorable charge−charge interactions, and it was expected that amino acid substitutions at these positions will be stabilizing. Thus, the increase in stability was *qualitatively* predicted on the basis of the change in the energy of charge−charge interactions (Table 1 and Figure 3).

Calculations of energies of charge−charge interactions by the TK−BK procedure also predicted that the positive charge on Arg72 has overall unfavorable contribution, and thus, substitution of the basic side chain should have a stabilizing effect. However, the R72Q variant of ubiquitin at pH 5.0 is 1.4 kJ/mol less stable than the wild type. Arg72 is the only residue among the studied ones that forms additional interactions. Analysis of the three-dimensional structure of the wild-type ubiquitin according to the algorithm of Stickle et al. (*23*) shows that NE of Arg72 can possibly form hydrogen bonding interactions with the backbone carboxyl oxygen of Glu41 situated 3.01 Å away. Thus, amino acid substitution of Arg72 with Gln not only will affect charge−charge interactions in the ubiquitin molecule but also will lead to a loss of one hydrogen bond. The energetic contribution of hydrogen bonding to protein stability is well established and is estimated to be on the order of 4 kJ/mol (*4, 24*). Taking this into account, we should expect that the combined effect of charge−charge interactions and the loss of hydrogen bonds upon amino acid substitution at position 72 should lead to a decrease in stability, exactly what we observe experimentally. We believe that the results obtained for the R72Q variant of ubiquitin are very important. They show that charge−charge interactions are not the only contributors and that quantitative prediction of protein stability is much more complex and requires accounting for other interactions.

Nevertheless, the data obtained in this study provide experimental evidence for the feasibility of increasing protein thermostability by improving charge−charge interactions on the surface of the protein. It is also possible that optimized surface charge−charge interactions provide structural determinants for enhanced stability of proteins from thermophilic organisms. This idea is not far from the hypothesis proposed by Max Perutz in 1978 who wrote "Enzymes of thermophile bacteria owe their extra stability mostly to additional salt bridges" (*25*). Detailed comparative analysis of electrostatic interactions to the stability of homologous proteins from mesophilic and thermophilic organisms also suggests that not only the number of salt bridges but also their relative location is important (*26, 27*). Despite the relative success in the stabilization of the ubiquitin molecule via improved charge−charge interactions, more experimental and theoretical studies exposing strengths and weaknesses of this approach are necessary.

REFERENCES

1. Perry, K. M., Onuffer, J. J., Gittelman, M. S., Barmat, L., and Matthews, C. R. (1989) *Biochemistry 28*, 7961−7968.
2. Serrano, L., Kellis, J. T., Jr., Cann, P., Matouschek, A., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 783−804.

3. Scholtz, J. M., and Baldwin, R. L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 95−118.

4. Pace, C. N. (1995) *Methods Enzymol. 259*, 538−554.

5. Matthews, B. W. (1995) *Adv. Protein Chem. 46*, 249−278.

6. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T., and DeGrado, W. F. (1995) *Science 270*, 935−941.

7. Smith, C. K., and Regan, L. (1995) *Science 270*, 980−982.

8. Cordes, M. H., Davidson, A. R., and Sauer, R. T. (1996) *Curr. Opin. Struct. Biol. 6*, 3−10.

9. Baldwin, R. L., and Rose, G. D. (1999) *Trends Biochem. Sci. 24*, 26−33.

10. Baldwin, R. L., and Rose, G. D. (1999) *Trends Biochem. Sci. 24*, 77−83.

11. Ramos, C. H., Kay, M. S., and Baldwin, R. L. (1999) *Biochemistry 38*, 9783−9790.

12. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M. P., Thurlkill, R. L., Scholtz, J. M., and Pace, C. N. (1999) *Protein Sci. 8*, 1843−1849.

13. Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I., and Sanchez-Ruiz, J. M. (1999) *Biochemistry 38*, 8138−8149.

14. Tanford, C., and Kirkwood, J. G. (1957) *J. Am. Chem. Soc. 79*, 5333−5339.

15. Bashford, D., and Karplus, M. (1991) *J. Phys. Chem. 95*, 9556−9561.

16. Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., Vierstra, R. D., Hatfield, P. M., and Cook, W. J. (1987) *J. Biol. Chem. 262*, 6396−6399.

17. Makhatadze, G. I., Lopez, M. M., Richardson, J. M., III, and Thomas, S. T. (1998) *Protein Sci. 7*, 689−697.

18. Ibarra-Molero, B., Makhatadze, G. I., and Sanchez-Ruiz, J. M. (1999) *Biochim. Biophys. Acta 1429*, 384−390.

19. Briggs, M. S., and Roder, H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2017−2021.

20. Rashin, A. A., and Honig, B. (1984) *J. Mol. Biol. 173*, 515−521.

21. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.

22. Makhatadze, G. I. (1999) *J. Phys. Chem. B 103*, 4781−4785.

23. Stickle, D. F., Presta, L. G., Dill, K. A., and Rose, G. D. (1992) *J. Mol. Biol. 226*, 1143−1159.

24. Myers, J. K., and Pace, C. N. (1996) *Biophys. J. 71*, 2033−2039.

25. Perutz, M. F. (1978) *Science 201*, 1187−1191.

26. Jaenicke, R. (1999) *Prog. Biophys. Mol. Biol. 71*, 155−241.

27. Xiao, L., and Honig, B. (1999) *J. Mol. Biol. 289*, 1435−1444.

28. Makhatadze, G. I., Clore, G. M., and Gronenborn, A. M. (1995) *Nat. Struct. Biol. 2*, 852−855.